

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2002, 01:52:40 ; Search time 65.99 seconds  
(without alignments)  
954.238 Million cell updates/sec

Title: US-10-112-331-4  
Perfect score: 1877  
Sequence: 1 LRALVFNHNLQVAEIPKSEI.....RLDAFRAITNDKNGENGP 364

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP\_ARCHAEA:.\*  
2: SP\_BACTERIA:.\*  
3: SP\_FUNGI:.\*  
4: SP\_HUMAN:.\*  
5: SP\_INVERTEBRATE:.\*  
6: SP\_MAMMAL:.\*  
7: SP\_MHC:.\*  
8: SP\_ORGANELLE:.\*  
9: SP\_PHAGE:.\*  
10: SP\_PLANT:.\*  
11: SP RODENT:.\*  
12: SP\_VIRUS:.\*  
13: SP\_VERTEBRATE:.\*  
14: SP\_UNCLASSIFIED:.\*  
15: SP\_VIRUS:.\*  
16: SP\_BACTERIAP:.\*  
17: SP\_ARCHAEP:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1532	81.6	364	1 Q9HHB5	Q9HHB5 PYROCOCUS
2	1494	79.6	364	1 P74630	P74630 SYNECHOCYST
3	1494	79.6	364	1 P74630	P74630 SYNECHOCYST
4	1494	79.6	364	1 P74630	P74630 SYNECHOCYST
5	147	7.8	527	16 Q97G73	Q97G73 CLOSTRIDIUM
6	142.5	7.6	443	17 Q97J70	Q97J70 SULLIOLBUS
7	136.5	7.3	895	17 Q97J70	Q97J70 SULLIOLBUS
8	130.5	7.0	560	17 Q97J70	Q97J70 SULLIOLBUS
9	128	6.8	923	16 Q9V294	Q9V294 PYROCOCUS
10	124	6.6	1362	17 Q9V294	Q9V294 PYROCOCUS
11	122.5	6.5	357	17 Q9H191	Q9H191 THERMOPHASM
12	120.5	6.4	324	17 Q9V294	Q9V294 PYROCOCUS
13	119	6.3	602	17 Q9V0M7	Q9V0M7 PYROCOCUS
14	111	5.9	526	16 Q93377	Q93377 TREPONEMA P
15	109.5	5.8	447	17 Q97ZD2	Q97ZD2 SULLIOLBUS
16	106	5.6	526	16 Q53278	Q53278 MYCOBACTERI

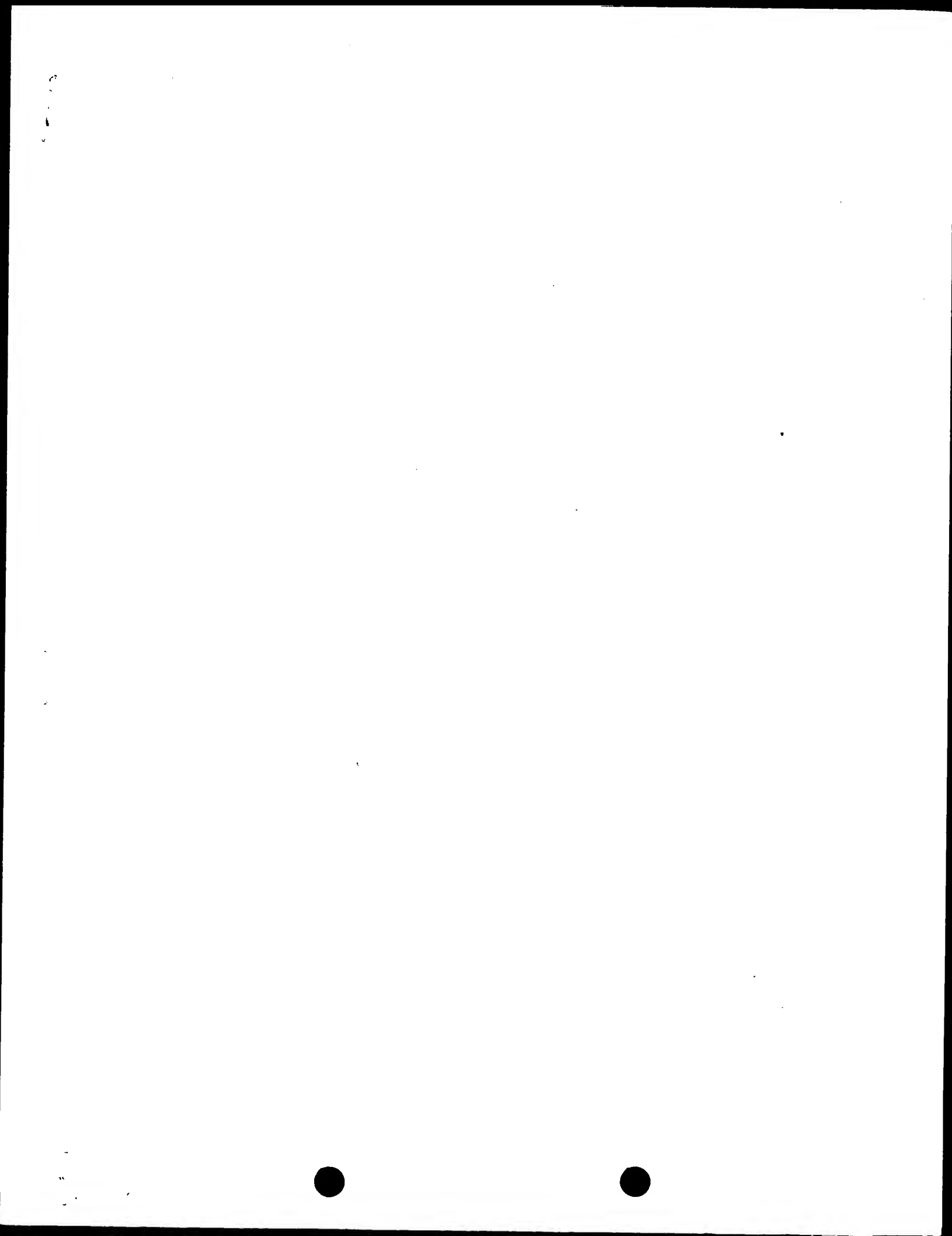
17	105.5	5.6	902	17 Q97YX0	Q97YX0 SULLIOLBUS
18	104	5.5	704	10 Q9ZVN2	Q9ZVN2 ARABIDOPSIS
19	103.5	5.5	458	10 Q9MBD0	Q9MBD0 PYRUS PYRIF
20	103.5	5.5	701	2 Q93MG7	Q93MG7 THIOBACILLU
21	103	5.5	754	13 Q13131	Q13131 ONCORHYNCHU
22	102	5.4	754	13 Q13132	Q13132 ONCORHYNCHU
23	100.5	5.4	266	16 Q97M42	Q97M42 CLOSTRIDIUM
24	100.5	5.4	619	16 Q83182	Q83182 TREPONEMA P
25	99	5.3	4845	11 Q88738	Q88738 MUS MUSCULU
26	98	5.2	503	5 Q17585	Q17585 CAENORHABDI
27	98	5.2	764	16 Q67347	Q67347 AQUIFEX AEO
28	97.5	5.2	402	2 Q68779	Q68779 YEASTINIA PE
29	97.5	5.2	605	16 Q9ZDR2	Q9ZDR2 RICKETTSIA
30	97.5	5.2	627	10 Q9ZDR1	Q9ZDR1 ARABIDOPSIS
31	97.5	5.2	684	10 Q9M254	Q9M254 ARABIDOPSIS
32	97.5	5.2	888	16 Q98P02	Q98P02 MYCOPLASMA
33	97.5	5.2	1085	16 Q25577	Q25577 HELICOBACTE
34	97.5	5.2	510	16 Q66973	Q66973 AQUIFEX AEO
35	97	5.2	531	17 Q97BF8	Q97BF8 THERMOPHASM
36	97	5.2	610	2 Q32582	Q32582 ESCHERICHIA
37	97	5.2	929	5 Q9YR32	Q9YR32 DROSOPHILA
38	96.5	5.1	197	13 Q918E8	Q918E8 FRUGO RUBR
39	96	5.1	636	2 Q87139	Q87139 VIBRIO CHOL
40	96	5.1	636	2 Q34235	Q34235 VIBRIO CHOL
41	96	5.1	1150	16 Q9YU58	Q9YU58 STAPHYLOCO
42	95.5	5.1	1134	17 Q9YU58	Q9YU58 PYROCOCUS
43	95.5	5.1	1326	5 Q16928	Q16928 CAENORHABDI
44	95	5.1	296	2 Q9Z597	Q9Z597 STREPTOMYCE
45	95	5.1	394	17 Q90YCO	Q90YCO PYROCOCUS

ALIGNMENTS

RESULT ID	Q9HHB5	PRELIMINARY	PRT	364 AA.
AC	Q9HHB5	01-MAR-2001 (TREMBLER, 16, Created)		
DT	01-MAR-2001 (TREMBLER, 16, Last sequence update)			
DT	01-MAR-2001 (TREMBLER, 19, Last annotation update)			
DE	ALPHA-GALACTOSIDASE.			
GN	GAL.			
OS	Pyrococcus furiosus.			
OC	Archaea, Euryarchaeota; Thermococcales; Pyrococcus.			
OX	NCBI_TaxID=2261.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=DSM 3638.			
RA	Verhees C.H.;			
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF195244; AAC28455.1;			
DR	InterPro: IPR004300; Glyco_hydro_57.			
DR	Pfam: PF03065; Glyco_hydro_57; 1			
DR	SEQUENCE 364 AA; 41545 MW; 3E8E0AF5BDCCF2A5 CRC64;			

Query Match: 81.6%; Score 1532; DB 1; Length 364;  
Best Local Similarity: 79.1%; Pred. No. 4.9e-117;  
Matches 287; Conservative 36; Mismatches 38; Indels 0; Gaps 0;

QY	1	LRALVFNHNLQVAEIPKSEIPKVIKAYIPVETLIKELIPGINTGYTLKPKRID 60
DB	1	MRALVFHNLQVAEIPKSEIPKVIKAYIPVETLIKELIPGINTGYTLKPKRID 60
QY	61	LVKGIASDLLEITSTTHAIPPLPSRYEAQVQRORREKVELFELSPGFMLPELAY 120
DB	61	LVKGIASDLLEITSTTHAIPPLPSRYEAQVQRORREKVELFELSPGFMLPELAY 120
QY	121	DLITPAIKDKNGEYLFADGEAMLFSAHNSAIKIPKPLPPLIKAKREKRRYISYLLG 180
DB	121	DLITPAIKDKNGEYLFADGEAMLFSAHNSAIKIPKPLPPLIKAKREKRRYISYLLG 180



Query Match 79.6%; Score 1494; DB 17; Length 364;  
Best Local Similarity 75.8%; Pred. No. 6.3e-114;  
Matches 275; Conservative 49; Mismatches 39; Indels 0; Gaps 0;

DB 181 LRELKRAIKLVEGKVTLLKAVKDIAVPVAVNTAVMLGIRLPLMDPKKVASWIEDKO 240  
181 LRELKRAIKLVEGKVTLLKAVKDIAVPVAVNTAVMLGIRLPLMDPKKVASWIEDKO 240

DB 241 NLLYGTDFIEFGYRDIAGYRMSVBEGLLEVIDELNSLCLPSLKHSGELVLTSSMAP 300  
241 ELLYGTDFIEFGYRDIAGYRMSVBEGLLEVIDELNSLCLPSLKHSGELVLTSSMAP 300

DB 301 DKSLLRWTEDEGNARLNLSTYMDGELAFLENSDARGMEPLPERRLDAFRAIYNDWGE 360  
301 DKSLLRWTEDEGNARLNLSTYMDGELAFLENSDARGMEPLPERRLDAFRAIYNDWGE 360

DB 361 NGE 363  
361 NGE 363

DB 361 NGE 363

RESULT 2  
PRELIMINARY: PRT: 364 AA.

AC 058106; PRELIMINARY: PRT: 364 AA.  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 41.8 KDA PROTEIN PH0368.  
GN PH0368.  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
NCBI\_TaxID=53953;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-073;  
RX MEDLINE=98344137; PubMed=9679194;  
RA Kawarayashi Y., Sawada M., Horikawa H., Halkawa Y., Hino Y.,  
Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
Funahashi T., Tanaka T., Kudoh Y., Yamazaki Y., Kushiida N., Oguchi A.,  
Masuchi Y., Shizuya H., Kikuchi H.;  
\*Complete sequence and gene organization of the genome of a hyper-  
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
RT DNA Res. 5:55-76(1998)  
DR EMBL: AF000002; BAA29442.1;  
DR InterPro: IPR004300; Glyco\_hydro\_57.  
DR Pfam: PF03065; Glyco\_hydro\_57; 1.  
KM Hypothetical protein: Complete proteome.  
SQ SEQUENCE 364 AA; 41755 MW; 7B4B36A4A975BAD CRC64;

Query Match 8.7%; Score 162.5; DB 16; Length 529;  
Best Local Similarity 21.4%; Pred. No. 7.1e-05;  
Matches 101; Conservative 47; Mismatches 140; Indels 183; Gaps 19;

DB 301 DKSLLRWTEDEGNARLNLSTYMDGELAFLENSDARGMEPLPERRLDAFRAIYNDWGE 360  
301 DKSLLRWTEDEGNARLNLSTYMDGELAFLENSDARGMEPLPERRLDAFRAIYNDWGE 360

DB 361 NGE 363  
361 NGE 363

DB 361 NGE 363

RESULT 3  
PRELIMINARY: PRT: 529 AA.

AC P74630; PRELIMINARY: PRT: 529 AA.  
DT 01-FEB-1997 (TREMBLrel. 02, Created)  
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE HYPOTHETICAL 62.1 KDA PROTEIN.  
GN SL0735.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97061201; PubMed=8905231;  
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
Miyajima N., Hirosewa M., Sugijara M., Sasamoto S., Kimura T.,  
Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
Tabata S.;  
\*Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
DR EMBL: D90916; BAA18743.1;  
DR InterPro: IPR003803; DUF200.  
DR Pfam: PF02651; DUF200; 1.  
KM Hypothetical protein: Complete proteome.  
SQ SEQUENCE 529 AA; 62075 MW; 4C1A45048A784E30 CRC64;

1



## FEATURES

Location/Qualifiers  
Source

1. 221284  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-312A14"

misc\_feature  
1. 1199  
/note="assembly\_name:Contig9"  
misc\_feature  
1300. 2874  
/note="assembly\_name:Contig10"  
misc\_feature  
2975. 4657  
/note="assembly\_name:Contig11"  
misc\_feature  
4758. 6622  
/note="assembly\_name:Contig12"  
misc\_feature  
6723. 10907  
/note="assembly\_name:Contig13"  
misc\_feature  
11008. 14713  
/note="assembly\_name:Contig14"  
misc\_feature  
14814. 20358  
/note="assembly\_name:Contig15"  
misc\_feature  
20459. 28691  
/note="assembly\_name:Contig16"  
misc\_feature  
28792. 42633  
/note="assembly\_name:Contig17"  
misc\_feature  
42734. 70133  
/note="assembly\_name:Contig18"  
misc\_feature  
70234. 106104  
/note="assembly\_name:Contig19"  
misc\_feature  
106205. 142909  
/note="assembly\_name:Contig20"

misc\_feature  
143010. 179097  
/note="assembly\_name:Contig21"  
misc\_feature  
179198. 221284  
/note="assembly\_name:Contig22"  
misc\_feature  
/note="assembly\_name:Contig20"  
vector\_side:left"  
clone\_end:SP6  
misc\_feature  
143010. 179097  
/note="assembly\_name:Contig21"  
misc\_feature  
179198. 221284  
/note="assembly\_name:Contig22"  
misc\_feature  
/note="assembly\_name:Contig20"

BASE COUNT 71097 a 38619 c 39582 g 70672 t 1314 others  
ORIGIN

Query Match 1.9% Score 21: DB 2: Length 221284;  
Best Local Similarity 100.0% Pred. No. 2.8;  
Matches 21: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 321 gcttcctccaaaggattctg 341  
|||||  
Db 15968 GCTTCTCCAAAGGATTCTG 15988

RESULT 14  
AR076196 AR076196 52 bp DNA linear PAT 30-AUG-2000  
LOCUS AR076196  
DEFINITION Sequence 1 from patent US 5958751.  
ACCESSION AR076196  
VERSION AR076196.1 GI:10002942  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 52)  
AUTHORS Murphy, D. and Reid, J.  
TITLE .alpha.-galactosidase  
JOURNAL Patent: US 5958751-A 1 28-SEP-1999;  
FEATURES  
source  
1. 52  
/organism="unknown"  
BASE COUNT 18 a 10 c 11 g 13 t  
ORIGIN

Query Match 1.8% Score 20: DB 6: Length 52;  
Best Local Similarity 100.0% Pred. No. 16;  
Matches 20: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 2 tggagcgctgcgtcttcac 21  
|||||  
Db 33 TGAGAGCGCTGCTTTCAC 52

RESULT 15  
TBILTA124/c  
LOCUS TBILTA124 1688 bp mRNA linear INV 15-NOV-1991  
DEFINITION T.brucei mRNA for variant surface protein ILTat 1.24.  
ACCESSION X56767  
KEYWORDS variant surface glycoprotein.  
SOURCE Trypanosoma brucei.  
ORGANISM Trypanosoma brucei  
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;  
Trypanosoma.

REFERENCE  
AUTHORS Carrington, M.  
TITLE Direct Submission  
JOURNAL Submitted (20-NOV-1990) M. Carrington, CAMBRIDGE UNIVERSITY,  
DEPARTMENT OF BIOCHEMISTRY, TENNIS COURT ROAD, CAMBRIDGE CB2 1QW,  
UK

REFERENCE  
AUTHORS Carrington, M., Miller, N., Blum, M., Roditi, I., Wiley, D. and  
Turner, M.  
TITLE Variant specific glycoprotein of Trypanosoma brucei consists of two  
domains each having an independently conserved pattern of cysteine  
residues  
J Mol. Biol. 221 (3), 823-835 (1991)

JOURNAL MEDLINE  
FEATURES  
source  
Location/Qualifiers  
1. 1688  
/organism="Trypanosoma brucei"  
/strain="ssp. brucei"  
/isolate="M1G 209"  
/db\_xref="taxon:5691"  
1. 18  
/note="mini exon"  
misc\_feature  
52. 1596  
/codon\_start=1  
/product="variant surface glycoprotein ILTat 1.24"  
/protein\_id="CAA0086.1"  
/db\_xref="GI:10454"  
CDS  
52. 1596  
/db\_xref="SWISS-PROT:P26329"  
/translation="MVRNLIQSLVLLVLLIVETNRGCVYELMOPCELTAEIR  
KTAGVAKKVNDSLSNPKTLTKMKLTPAKPESKEALTLRALALANTDLRAR  
DNANGIDRAVRATVASEAGALFSGIQLTHDSTDYTCLSASGGSGNGNANASO  
GCKPLALPELTEDSYNTDVISDKGPKISPLTNAOOGSGEGGLQVAGAGATY  
GVORSGSRINILGAIIVASAQOOPTRDLSDFGTARNOADTLTGRAHASTIEL  
AQGRPKQTEVEYTKLILAAOKTRALDLSIKFOLASTGKTSYKEDENIKTEYFGKT  
NIEALMNVKEEKYKADPEDPSKESKISDLNTEEOLOAVYAVATMLAKQARD  
IAKLETEIADRGKSPENECNKITEEPKSEKIKCSHAKYKAEKNCORNSYASAS  
GVPTQVOTAGADTTAERCKGKGRDKSPCKWEGGTCCKDSLANKORALSVASAA  
FVALIF"

sig\_peptide 52. 120  
mat\_peptide 121. 1593  
BASE COUNT 594 a 389 c 389 g 316 t  
ORIGIN

Query Match 1.8% Score 20: DB 3: Length 1688;  
Best Local Similarity 100.0% Pred. No. 14;  
Matches 20: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 336 attcgctccagagctcg 355  
|||||  
Db 160 ATTCTGCTGCCAGACTCG 141

Search completed: June 21, 2002, 23:04:10

6  
3



1  
5  
2

Mon Jun 24 08:37:17 2002

Job time: 7436 sec

us-10-112-331-3.rge









```

/product-"putative transporter"
/protein_id-"AAG28456.1"
/db_xref="GI:11066094"
/translation-"MGNIRRVSIYVLLIIMAAFLMADNOLLPNTQOIKERGIGETO
MGLVSIKATSAIILIIIMGLADIKORRKLIVGILGICFICFLTAIVQYTWOLAM
RFLTGIGISIIIPICYSIADMPFEERKRGSIYETAFGFTGLFMIIAGNIMASRT
PRTIAAPRFLIAPLFIYIAEPRKRGKRGKRIALIEKYEYLRNLEAIKRSQTK
NIIILFLOGIIIGTVPWGIILVWVSFLQVTRGMDSATFVLLIIGISVFSGLGFE
VGDYENARRRGGRATITGIAIFLGLMGLASIGLILYPSLSTVHNGIATISILFIOF
VSYAPNRAIVSOVNPEDRGVFLGFLIDNKGKATIGLPGFLIETKSMGMTEP
LAYOTLLIGALFIPCALWIMIKRSPEDDKYKEILKRAEELISKPT"
complement(3457..4722)
/gene="bglA"
/complement(3457..4722)
/gene="bglA"
/feature="family 1 glycosyl hydrolase"
/codon_start=1
/translation-"beta-glucosidase"
/product="beta-glucosidase"
/protein_id-"AAG28457.1"
/db_xref="GI:11066095"
/translation-"MPLKPEEFLFGTATAAHOIEGDKNMDMWWYEOIGKLPYKSG
ACNHEFEKEDIOIAMSAGYNAVRSISMSPLPEENKFNENAFRNYEDIIDLLAN
IPLVTLHHTSPIMPMKRGFLRENNKPEKYEYAELEKRLIATNEPWWYV
EMGYLTATWPPPIKSPFAKFAVSNLAAHLLAVEILHGKQVGIKVPMLPATDK
ERDKAAERADNLEWYFLDAIMSGVIRGARAVYPOSDDFIGINVTASVESHV
NPLKFFEDAKLADVSEKRTOMGWSYPRGIIYALKASKYGPILYITENGIALTDEW
RIEFLIOHLOVYHKAIEDLDVGRGYFYSYMNEMRGFEPRGLVVDYETFERPR
RKSATYIGSIKSKIDELIEKYLSSL"
4779..5762
/gene="galT"
/feature="galT"
/codon_start=1
/translation-"putative galactose-1-phosphate
uridylyltransferase"
/protein_id-"AAG28458.1"
/db_xref="GI:11066096"
/translation-"MRELRYNPITGOWIMVSAVRKRRPMPRTNCPGSGSEETGYG
KVLILPNRPMISFADPESEFEYKRAKRAIGOSVIEPTDHLKDLDEPRDMYK
VVELWVENTKTLKRNPHAYALAFRNKGESEISLTHRGOLYALPTIPLKRLTEN
SRKYKSKNECLPRILKEKESERIIYENDNTVPEPFASMPFEMHITPKRHVOY
LTDANKESIESLAIOITGTITLIEROMPVTMMIFQAPFKNYKRYHIEFYPI
LRDNGKIKRYAAGISMGTEWETFYDAIPEENARKLEACKRVYSQRILNKGCIT"
BASE COUNT 1896 a 1123 c 1174 g 1569 t
ORIGIN

```

Query Match 54.6% Score 598; DB 1; Length 5762;  
 Best Local Similarity 71.7% Pred. No. 8.6e-167;  
 Matches 784; Conservative 0; Mismatches 310; Indels 0; Gaps 0;

```

OY 2 tgaagagcgcgtcttcaagcaacccagatgacgaatcccaagagcgaatcc 61
DB 1057 TGAGAGCACTAGTCTTTCATGGGAATTACATATGACGAATACCAAGCAAAATTC 1116
OY 62 caaagtcataagagacacacacacacacacacacacacacacacacacacac 121
DB 1117 CAAAGGTTATGAAACATACACTTCCCTATTTTTCAGAACTCTATGAGAGAGATAC 1176
OY 122 ctttgagctaaacacacacacacacacacacacacacacacacacacacac 181
DB 1177 CTTTGGCTTAAACATACACGAGATACATCAATGCTTCTCCCAAAAGATTGATTGCC 1236
OY 182 tcttaagagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 241
DB 1237 TCATTAAGAGAGGAAATAGAGAGTGGCCCATATGAGATTTCTTGGAACAAGATTATCTCAGC 1296
OY 242 caatactccctccctccctccctccctccctccctccctccctccctccctcc 301
DB 1297 CAATCTCCCACTCTCTCCCACTCTCTAGAGTTGAGGCAATAAATAAAGAGATGAGAG 1356

```

```

OY 302 ttaaggaagccttcagagcttcccaagagattcggcgcagagcgcctatg 361
DB 1357 TGAAGAAGAAACATATTGGAAGTCCCCAGAGAGATTTCGCTCCCGAATTGGCATATG 1416
OY 362 acccgaataccctccacacacacacacacacacacacacacacacacacacac 421
DB 1417 ACCCAATAATTCACAGAAATTTGAGGACACACATATGATATCTCTTCGTCGAGGGG 1476
OY 422 aggcagctcttctcagctcactcactcagcagcagcagcagcagcagcagc 481
DB 1477 AGGCTATGCTATCTCTCAACATCTTCACTGCAATTAACCAATTAACCTTATACC 1536
OY 482 ccaaccttaagagcccaaggaaggaagcgttcagtagtaactcactccttcg 541
DB 1537 CTCATCTATTAAGAGCCCAAGAGGAGGAGGCTTACTGTACTTAATTAATCTTTTNGGCC 1596
OY 542 tcaaggaagccttgaagagcagcagcagcagcagcagcagcagcagcagcag 601
DB 1597 TTAGAGAGCTCAAAAAGGCGCATTAATCTTCTTTAAGAAAGACACATGAGAGAG 1656
OY 602 tcaagagacacagcagcagcagcagcagcagcagcagcagcagcagcagcag 661
DB 1657 TCAGAGAAATCGAAGCTATCCGCGTATGGTTTCCATCAATCTGCAATGATGCTGGAG 1716
OY 662 tcggaagcctcctctatgaactcctaagaagtgccgagcgtgagtagagcagagaca 721
DB 1717 CGGGAAGATTTCACCTATTGAAATCTTAAGAAAGTCCCAAAATGGGTAAAGAAAAAGATG 1776
OY 722 acattctctatacgcagcagcagcagcagcagcagcagcagcagcagcagcagcag 781
DB 1777 AATTTTCTTATGGAACAGACATAGAGTTCCTCGAGATACAGATATAGCAGGATATA 1836
OY 782 gaatgagtgtagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 841
DB 1837 AGATTACATTTTCCAACTATTGAGATTATTAAGAGCTTAAAGAGATTATAGGCTTAC 1896
OY 842 cctcagagcgtgacagcagcagcagcagcagcagcagcagcagcagcagcagcag 901
DB 1897 CCAGGAAGATTAAGCATACTGTAAGAAAGCTTATCTTAAGAAAGCTTAAAGTGGCTCCGG 1956
OY 902 ataagagccttgaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 961
DB 1957 ACNAAAGTTTGAGATTGACAGAGATGAGAGGAATGCAAGATTAAACATGCTTACT 2016
OY 962 acaatataagagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1021
DB 2017 CCTATATGATGAGAGACTGCTTATTTTATGCTGTAAGAAATGATGATCCCGATGTTGGAC 2076
OY 1022 cctccctgaagagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1081
DB 2077 CTCCTCCGGAAGAGAGATTGACGATTTAAAGCTATATACACCATTTGAGAGAGTGA 2136
OY 1082 atggaggaacctag 1095
DB 2137 ATGGGAACATTAG 2150

```

RESULT 3  
 AP000002/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Pyrococcus horikoshii (strain:OT3) DNA.  
 Pyrococcus horikoshii  
 Archaea: Euryarchaeota: Thermococci: Thermococcales:  
 Thermococcaceae: Pyrococcus.  
 Kawaiibayashi, Y., Sawada, M., Horikawa, H., Halkawa, Y., Hino, Y.,  
 Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y.,









